A SURVEY OF FACE RECOGNITION ON FEATURE EXTRACTION PROCESS OF DIMENSIONALITY REDUCTION TECHNIQUES

S.CHITRA, Dr.G.BALAKRISHNAN
1 Asstt Prof., Department of Computer Applications, SASTRA University, Tanjore
2 Director, Indra Ganesan College of Engineering, Tiruchirappalli
E-mail: 1chithismca@gmail.com , 2balakrishnan.g@gmail.com

ABSTRACT

Face Recognition is one of the most successful challenging applications in the field of computer vision and pattern recognition. Generally there are two types of recognitions such as intrusive recognition means that the user aware about the recognition i.e., Palm print recognition; the users have to place their palm in the scanner, whereas as face recognition is non-intrusive, without user cooperation it can able to recognize the person as authenticated person or not. The applications of face recognition are time attendance system, visitor management system and access control system, etc. the face recognition gives efficient performance under the controlled environment. But still we have the unsolved problems in real time applications. The dimensionality reduction is a most important task in the field of face recognition. In this paper, it proposed all the recent emerging techniques of feature extraction process in the dimensionality reduction.

Keywords: Face Recognition, Dimensionality Reduction, Feature Extraction, Feature Selection, Linear Methods, And Non Linear Methods.

1. INTRODUCTION

Face recognition is the most challenging work for the research persons from the year of 1990’s. The researchers gave satisfactory results for the still images i.e., images are taken under the controlled conditions. If the image contain the problems like illumination, pose variation, aging, hair inclusion then the performance of the recognition process leads to poor. Most of the researchers are concentrating on the real time applications. Many surveys are carried out on the topic of face recognition [1-6] they specify various existing techniques for feature extraction and the face recognition process.

Generally face recognition is classified as the process of face detection, feature extraction and face recognition. Image preprocessing work as removing the background details and normalize the image by rotation, scaling, resizing of the original image is carried out before the face detection process. The face detection is to detect the face from the normalized image, then the feature extraction process is used to extract the features from the detected face and finally the face recognition process is to recognize the face compared with face database which is already stored [1-6]. Figure 1 denotes the process of face recognition.

![Figure 1: process of face recognition system](image)

It is impossible to directly deal with raw data while the amount of data is increased. Dimension reduction is the task to solve the above problem by extracting the structured data and
remove the redundant data. If the training images is increased then the matrix of image also increased then it is called as a problem of “Curse of dimensionality” which is solved by dimensionality reduction techniques [7] states that there are two types of dimensionality reduction techniques as linear and non linear dimensionality reduction. The linear dimensionality reduction techniques are PCA, LDA, LPP, etc. and the non linear dimensionality techniques are ISOMAP, LLE, and so on.

The aim of this paper is to give emerging techniques for the dimensionality reduction in linear as well as non linear techniques. It can be arranged as follows, section 2 contains the information about the dimensionality reduction. section 3 have the details about literature review of dimensionality reduction, section 4 denotes the various techniques in the dimensionality reduction, section 5 denotes the summary of techniques in dimensionality reduction and section 6 contain the conclusion of this paper.

2. DIMENSIONALITY REDUCTION TECHNIQUES

2.1 Overview:

The most important problem in face recognition is the curse of dimensionality problem. The methods are applied to reduce the dimension of the studied space. When the system starts to memorizes the high dimensional data then it causes over fitting problem and also computational complexity becomes the heavy task. This curse of dimensionality problem is reduced by dimensionality reduction techniques [7].

In statistics, dimension reduction is the process of reducing the number of random variables under consideration \( R^N \rightarrow R^M \) (\( M<N \)), and can be divided into feature selection and feature extraction [8]. The basic flow of dimension reduction in face recognition is illustrated in figure 2.

The author [9] says that the various techniques are existed for solving the problem of curse of dimensionality. Out of those techniques some are linear methods and others are nonlinear. Linear methods is to transform data from high dimensional subspace into low dimensional subspace by linear mapping but it fails to work in the non linear data structure where as non linear methods are easily work in the complex non linear data structure. Compared to linear methods, non linear methods are very efficient while processing the problematic image like hair inclusion, lighting condition and so on. Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Locality Preserving Projections (LPP) are some popular linear methods and nonlinear methods include Isometric Mapping (ISOMAP) & Locally Linear Embedding (LLE).

According to the author [8], Feature selection is to find a subset of the original variables. Two strategies are filter (e.g. information gain) and wrapper (e.g. genetic algorithm) approaches. It occurs sometimes that data analysis such as regression or classification can be done in the reduced space more accurately than in the original space. Feature extraction is applying a mapping of the multidimensional space into a space of fewer dimensions. This means that the original feature space is transformed by applying a linear transformation. The brief introduction of feature extraction techniques are illustrated in the next section.

2.2 Linear Feature Extraction Of Dimensionality Reduction Techniques:

Generally the face recognition process is divided into 3 regions such as Holistic method use the original image as an input for the face recognition system. The examples for holistic methods are PCA, LDA, and ICA and so on. In Feature based method, the local feature point such as eye, nose, and mouth are first extracted, then it will be send to the classifier. Finally a Hybrid method is used to recognize both the local feature and whole face region [1-6].

In Dimensionality reduction, Feature extraction is an important task to collect the set of features from an image. According to the author [10], Feature extraction/transformation is a process through which a new set of features is created. The feature transformation may be a linear or nonlinear combination of original features. This survey provides some of the important linear and non linear techniques are listed as follows.
2.2.1 Principal Component Analysis (PCA):

PCA is one of the popular techniques for both dimensionality reduction and face recognition since 1990’s. Eigenfaces [12] built on the PCA technique is introduced by M.A.Turk and A.P.Pentland. It is a holistic approach where the input image is directly used for the process. PCA algorithm can be used to find a subspace whose basis vectors correspond to the maximum variance directions in the original n dimensional space. PCA subspace can be used for presentation of data with minimum error in reconstruction of original data. More survey papers are provided the information for PCA techniques [1-6]. MPCA and KPCA are fully based on the PCA technique.

2.2.2 Linear Discriminant Analysis (LDA):

LDA is one of the most famous linear techniques for dimensionality reduction and data classification. The main goal of the LDA consists in finding a base of vectors providing the best discrimination among the classes, trying to maximize the between-class differences, minimizing the within-class ones by using scatter matrices. It also suffers from small sample size problem which exists in high dimensional pattern recognition task where number of available sample is smaller than dimensionality of the samples. D-LDA, R-LDA, and KDAA are variations of LDA. This technique also discussed in more survey papers [6, 7, 8, 9, 10, 11, 13, 14 and 15].

2.2.3 Singular Value Decomposition (SVD):

SVD is an important factor in the field of signal processing and statistics. It is the best linear dimensionality reduction technique based on the covariance matrix. The main aim is to reduce the dimension of the data by finding a few orthogonal linear combinations of the original variables with the largest variance [9]. Most of the researches are also used this technique for face recognition.

2.2.4 Independent Component Analysis (ICA):

ICA is a statistical and computational technique for enlightening the hidden factors that underlie sets or random variables, measurements, or signals. ICA is superficially related to principal component analysis and factors analysis. The ICA algorithm aims at finding S component as independent as possible so that the set of observed signals can be expressed as a linear combination of statistically independent components. It use cosine measures to perform the covariance matrix and also it is better than the PCA and LDA performance.

2.2.5 Locality Preserving Projections (LPP):

LPP can be seen as an alternative to Principal Component Analysis (PCA). When the high dimensional data lies on a low dimensional manifold embedded in the ambient space, the Locality Preserving Projections are obtained by finding the optimal linear approximations to the eigenfunctions of the Laplace Beltrami operator on the manifold. As a result, LPP shares many of the data representation properties of nonlinear techniques such as Laplacian Eigenmaps or Locally Linear Embedding [7].

2.2.6 Multi Dimensional Scaling (MDS):

Multidimensional Scaling (MDS) [Borg and Groenen, 1997] is a linear Model for dimensionality reduction. MDS generates low dimensional codes placing emphasis on preserving the pair wise distances between the data points. If the rows and the columns of the data matrix D both have mean zero, the projection produced by MDS will be the same as that produced by PCA. Thus, MDS is a linear Model for dimensionality reduction having the same limitations as PCA.

2.2.7 Partial Least Squares:

Partial least squares is a classical statistical learning method. It is widely used in chemo metrics and bioinformatics etc. In recent years, it is also applied in face recognition and human detection. It can avoid the small sample size problem in linear discriminant analysis (LDA). Therefore it is used as a alternative method of LDA.

2.3 Non Linear Feature Extraction Of Dimensionality Reduction Techniques

Non-linear methods can be broadly classified into two groups: a mapping (either from the high dimensional space to the low dimensional embedding or vice versa), it can be viewed as a preliminary feature extraction step and visualization is based on neighbor’s data such as distance measurements. Research on non-linear dimensionality reduction methods has been explored extensively in the last few years. In the following, a brief introduction to several non-linear dimensionality reduction techniques will be given.

2.3.1 Kernel Principle Component Analysis (KPCA):

Kernel PCA (KPCA) is the reformulation of traditional linear PCA in a high-dimensional space that is constructed using a kernel function. In recent years, the reformulation of linear techniques using the ‘kernel trick’ has led to the proposal of successful techniques such as kernel ridge regression and Support Vector Machines. Kernel PCA computes the principal eigenvectors of the kernel matrix, rather than those of the covariance matrix. The reformulation of traditional PCA in kernel space is straightforward, since a kernel matrix is similar to the in product of the data points in the high-dimensional space that is constructed.
using the kernel function. The application of PCA in kernel space provides Kernel PCA the property of constructing nonlinear mappings.

2.3.2 Isometric Mapping (Isomap):

Most of the linear methods do not take the neighboring data points into an account. ISOMAP is a technique that resolves this problem by attempting to preserve pair wise geodesic (or curvilinear) distances between data points. The approximation of geodesic distance is divided into two cases. For, neighboring points, Euclidean distance in the input space provides a good approximation to geodesic distance and faraway points, geodesic distance can be approximated by adding up a sequence of “short hops” between neighboring points. ISOMAP shares some advantages with PCA, LDA, and MDS, such as computational efficiency and asymptotic convergence guarantees, but with more flexibility to learn a broad class of nonlinear manifolds [7].

2.3.3 Locally Linear Embedding:

Locally linear embedding (LLE) is another approach which addresses the problem of nonlinear dimensionality reduction by computing low-dimensional, neighborhood preserving embedding of high-dimensional data. It is a technique that is similar to ISOMAP in that it also constructs a graph representation of the data points. It describes the local properties of the manifold around a data point \( x_i \) by writing the data point as a linear combination \( w \) of its \( k \) nearest neighbors \( x_{ij} \) and attempts to retain the reconstruction weights in the linear combinations as good as possible [16,17].

2.3.4 Laplacian Eigenmaps:

A closely related approach to locally linear embedding is Laplacian eigenmaps. Given \( t \) points in \( n \)-dimensional space, the Laplacian eigenmaps Method (LEM) starts by constructing a weighted graph with \( t \) nodes and a set of edges connecting neighboring points. Similar to LLE, the neighborhood graph can be constructed by finding the \( k \) nearest neighbors. The final objectives for both LEM and LLE have the same form and differ only in how the matrix is constructed [16].

2.3.5 Stochastic Neighbor Embedding:

Stochastic Neighbor Embedding (SNE) is a probabilistic approach that maps high dimensional data points into a low dimensional subspace in a way that preserves the relative distances to near neighbors. In SNE, similar objects in the high dimensional space will be put nearby in the low dimensional space, and dissimilar objects in the high dimensional space will usually be put far apart in the low dimensional space [17]. A Gaussian distribution centered on a point in the high dimensional space is used to define the probability distribution that the data point chooses other data points as its neighbors. SNE is superior to LLE in keeping the relative distances between every two data points.

2.3.6 Semi Definite Embedding (SDE):

Semi definite Embedding (SDE), can be seen as a variation of KPCA and an algorithm is based on semi definite programming. SDE learns a kernel matrix by maximizing the variance in feature space while preserving the distances and angles between nearest neighbors. It has several interesting properties: the main optimization is convex and guaranteed to preserve certain aspects of the local geometry; the method always yields a semi positive definite kernel matrix; the eigenspectrum of the kernel matrix provides an estimate of the underlying manifold’s dimensionality; also, the method does not rely on estimating geodesic distances between far away points on the manifold. This particular combination of advantages appears unique to SDE.

3. RECENT WORKS IN LINEAR AND NON LINEAR FEATURE EXTRACTION OF DIMENSIONALITY REDUCTION TECHNIQUES:

In [20], B2DPCA, dimensionality reduction algorithm operates independently along row and column directions in order to better preserve the neighborhood relationship and to generate distinctive feature sets. It generates an image covariance matrix and further optimizes it exploiting optimal project axes. Once optimal projection axes is calculated, the dimensionality of every image is reduced along its columns to generate new image sets. The newly generated image sets are subsequently treated as a fresh database and a latest image covariance matrix and optimal projection axes are evaluated. Finally, every new image is pre-multiplied. Hence, unlike traditional 2DPCA, a twofold approach is adopted in B2DPCA algorithm to reduce image dimensionality.

In [19], new discriminative color features method first derives compact color features from the new color model by reducing the dimensionality of the color component images of the RGBr color model and apply DCT for dimensionality reduction for the color component images of the RGBr color model. As its basis vectors are fixed, the DCT is able to significantly improve the computational efficiency for dimensionality reduction. The DCT
method transforms the color component images of the RGBr color model from the spatial domain to the frequency domain. As the low frequency features in the frequency domain display good information packing capability, they are selected to form low dimensional pattern vectors to represent the Rr, Gr, and Br color component images.

PLS is used to reduce the influence of pose in [21]. PLS models the relations between two sets of variables by means of score vectors. “The underlying assumption of all PLS methods is that the observed data is generated by a system or process which is driven by a small number of latent variables.”

Modular PCA [22] divides the face images into small patches and applies PCA on each set of patches. Modular LDA [22] uses a set of independently trained observers on different parts of faces. Each observer performs LDA independently by projecting faces to a lower dimensional subspace and performing recognition. The final result is achieved by using a simple sum rule on the recognition results. This approach gives efficient performance when illumination problem arises.

In [23], they construct a multi-directional orthogonal gradient phase face (OGPF) algorithm by introducing directional derivative into image calculation, which can provide more complete face feature description. In multi-directional OGPF algorithm, can be generated from one image, which extends the samples of each person and is beneficial to many subspace based dimensionality reduction techniques, such as classical and effective LDA. LDA cannot be directly performed on Gradient faces when there is only one sample of each person for training. Multi directional OGPF method is an effective algorithm for single sample face recognition under not only illumination but also expression, decorations, etc. Furthermore, multi-directional OGPF plus PCA&LDA algorithm enhances the discriminating power of original orthogonal gradient phase face and reduces the length of template.

The modification of LDA (FLDA) [24] is to introduce the fuzziness into the every projector vector of the classes. In this method a vector is assigned the membership grades for every class based on k-nearest neighbor. This Fuzzy k-nearest neighbor algorithm of is used to calculate the member ship grades of all the vectors.

2DNPP [25] is directly derived from the NPP algorithm. It is a two dimensional of NPP algorithm which comes under the framework of LLE. It does not require any parameter selection during the neighborhood weighting. The performance of this algorithm is better than the 2DLPP, LPP, NPP, 2DLDA, 2DPCA.

Block-wise 2D-KPCA/GDA is a non linear feature extraction of the dimensionality reduction which proposed by the author of [26]. This method will solves the drawbacks of 2DPCA/LDA by exploring the higher order statistics among the rows of input images rather than direct extension of 2DPCA/LDA to kernel induced feature space. It includes the concept of block manifolds to utilize the local characteristics of input space.

An incremental 2DLDA is proposed [27] to update the discriminant subspace instead of full re-training whenever a new training sample is added. A closed-form solution for updating the between-class scatter matrix and within-class scatter matrix using the new samples is derived. The advantages are solves the small sample size problem and is able to extract more discriminant information, the number of the classes or the chunk size can be very large because the memory cost for maintain the between-class and within-class scatter matrices are low.

In [28], they extend the method of 2DPCA and BDPCA to non-Euclidean space as an Laplacian BDPCA (LBDPCA) method and it is to improve the robustness of 2DPCA and BDPCA by defining the Laplacian row total scatter matrix and the Laplacian column total scatter matrix, calculating the eigenvectors of the scatter matrices and finally the image matrix was projected onto the projectors to extract the LBDPCA feature. LBDPCA considers the structural information of the image samples and the distribution information embedded in the original images.

In LPP, it suffers with the problem of small sample size problem, i.e., the eigen equation not able to solve directly where as the LPP subtraction transforms the objective function of LPP into a new function, which allows the eigen equation to solve directly [29]. A novel feature extraction method, namely [30], generalized two-dimensional FLD (G-2DFLD) method, which is based on the original 2D image matrix. The G-2DFLD algorithm maximizes class separability from both the row and column directions simultaneously, resulting in smaller image feature matrix. The image feature matrix is smaller than the 2DPCA and 2DFLD algorithms.
4. SUMMARY OF LINEAR AND NON LINEAR METHODS:

The figure 3 shows the various types of linear and non-linear feature extraction methods in the dimensionality reduction technique and table 1 shows the advantages and disadvantages of the existing methods for linear feature extraction.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>Correlation between components of the data vector be clearly viewed and it capture the second order correlation value. It uses the eigenface</td>
<td>Identifies the linear combinations of variables and Ignore the high order correlation value</td>
</tr>
<tr>
<td>LDA</td>
<td>Linear mapping, dimensionality of the subspace is limited by the number of classes of the data</td>
<td>Cannot handle data in which the individual classes are far from Gaussian, suffers from SmallSample Size problem</td>
</tr>
<tr>
<td>MDS</td>
<td>Best preserves the pairwise distances between every two data points</td>
<td>Same limitations as PCA</td>
</tr>
<tr>
<td>LPP</td>
<td>It shares many of the non linear properties for data representation. Gaussian weights are used to reduce the space</td>
<td>Occurrence of parameter sensitive.</td>
</tr>
<tr>
<td>SVD</td>
<td>Robust method of storing large images as smaller</td>
<td>Computation is hard and does not work for subsequence indexing.</td>
</tr>
<tr>
<td>PLS</td>
<td>It can avoid the small sample size problem and it also reduces the influence of pose.</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: linear methods of feature extraction
The table 2 refers to the nonlinear perspective of feature extraction will be illustrated as follows.

Table 2: Non Linear methods of Feature Extraction

<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISOMAP</td>
<td>Preserves a pair wise geodesic distances between data points</td>
<td>Unclearness of how to evaluate the maps on new test data.</td>
</tr>
<tr>
<td>KPCA</td>
<td>Computes the kernel matrix to reduce the dimension</td>
<td>Size of the kernel matrix is proportional to the square of the number of instances in the dataset.</td>
</tr>
<tr>
<td>LLE</td>
<td>Keep the intrinsic distribution of face sample data</td>
<td>Out of sample problem will be occur.</td>
</tr>
<tr>
<td>Laplacian Eigenmap</td>
<td>Successful in reducing the dimensionality for semi-supervised learning</td>
<td>Unclearness of how to evaluate the maps on new test data.</td>
</tr>
<tr>
<td>SOM</td>
<td>Very efficient even the image contain noise</td>
<td>If the number of classes increase then the chances of mismatch are more.</td>
</tr>
<tr>
<td>NPP</td>
<td>Data-driven weights are used to solve the least square problem</td>
<td>It needs to solve a generalized eigenvalue problem</td>
</tr>
</tbody>
</table>

5. CONCLUSION:

Generally Face Recognition can be divided into three parts they are Holistic approach, feature based approach and Hybrid based approach. The holistic approaches take the whole face image as a raw data and recognize the face. In feature based approach, the features of a face like eyes, nose and mouth are extracted and then recognize it where as the hybrid approach is the combination of both the Holistic and feature based methods. The survey paper concentrates on the approach of holistic based recognition. If we want to recognize the face we need three processes should be done they are face detection, dimensionality reduction and face recognition. The dimensionality reduction is used to solve the curse of dimensionality. It can be divided into two parts they are Feature Extraction and Feature Selection. The feature extraction process can be broadly classified into four types they are linear method, non linear methods, Multi linear methods and tensor space methods. Among those categories this paper collects the information about the various methods included in the linear and non linear feature extraction process.

PCA, LDA, ICA are the most well known linear feature extraction process for past more than 10 years where as KPCA, ISOMAP, LLE are the famous technique in non linear feature extraction. Now the researches are concentrated on combining both linear and non linear methods to reduce the dimensionality reduction and also for feature extraction methods. The contribution of this paper is to give details about the dimensionality reduction techniques in both linear and non linear and also how to deal with small sample size problem. There are so many techniques are available, even though still there is a problem while the occurrence of illumination and pose changes. The future work will concentrate on the above said problem.

REFERENCES:

[5] Xiaoyang Tana,b, Songcan Chena,c,*, Zhi-Hua Zhoub, Fuyan Zhangb,” Face recognition from a single image per person:Asurvey”,Published in Elseiver,2006


[17] Renqiang Min,” A Non-linear Dimensionality Reduction Method for Improving Nearest Neighbour Classification”,2005


[21] Annan Li a,b, Shiguang Shan a, Xilin Chen a, Wen Gao c,a,” Cross-pose face recognition based on partial least squares”, Aug 2011

[22] Amirhosein Nabatchian n, EsamAbdel-Raheem,MajidAhmadi,” Illumination invariant feature extraction and mutual-information-based local matching for face recognition under illumination variation and occlusion “,Mar-2011

[23] Xi Chen n, JiashuZhang ,” Illumination robust single sample face recognition using multi-directional orthogonal gradient phase faces “,Apr-2011


[26] Armin Eftekhari a, Mohamad Forouzanfar b,c,* Hamid Abrishami Moghaddam c,d,Javad Alirezaie e,” Block-wise 2D kernel PCA/LDA for face recognition”, June 2010


[28] Wankou Yang a,n, ChangyinSun a, LeiZhang b, KarlRicanek c,” Laplacian bidirectional PCA for face recognition”, Oct-2010

[29] Yong Xu a,Fengxi Song b, Ge Feng a, Yingnan Zhao c,” A novel local preserving projection scheme for use with face recognition”, 2010

AUTHOR PROFILES:

S.Chitra is a research student of Anna University, Trichy. Currently, she is an Assistant Professor- I at SASTRA University, Tanjore.

Dr.G.Balakrishnan received the PhD degree in computer science from the University of Malaysia Sabah, Malaysia in 2006. Currently, he is a Director of Indra Ganesan College of Engineering. His research interests include Image Processing.